

Video Analytics Towards Vision Zero



Traffic Video Analytics

Bellevue, Washington

Microsoft Corp.

Case Study Report

December 2019



Executive Summary

Traffic safety and traffic congestion represent one of the biggest problems in modern cities, and improving these is a major promise of smart cities of the future. Microsoft collaborated with the City of Bellevue to use video camera feeds for traffic video analytics. The partnership's vision was to use the widely deployed traffic camera feeds for traffic analytics on road users as opposed to traditional mechanical or manual approaches. A primary objective of the partnership was to evaluate if video analytics can produce accurate outputs on live video feeds to produce actionable insights to inform Vision Zero strategies. Bellevue is committed to implementing a Vision Zero Action Plan, with the goal of zero traffic deaths and serious injuries by 2030.

The Microsoft team developed a video analytics platform that analyzed videos to produce directional counts of traffic users (vehicles, bicycles, etc.). The results were aggregated into a video analytics dashboard, that was deployed at the City of Bellevue's Traffic Management Center from Jul 2017 to Nov 2018, and produced live alerts on abnormal traffic volumes. The dashboard also helped Bellevue transportation planners understand traffic patterns over long periods of time. For example, it provided the perspective on vehicle and bicycle patterns, prior to and after construction of a dedicated bicycle lane on 108th Avenue. Finally, the video analytics system's insights on directional volumes and unusual traffic patterns provided an additional tool for real-time traffic operations in the Bellevue Traffic Management Center.

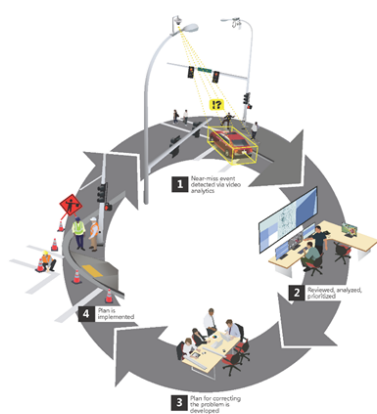
Key Outcomes

- The City of Bellevue's Traffic Management Center deployed the dashboard of directional traffic volumes that provided live data as well as alerts based on historical volumes.
- Evaluation of the bike lane on 108th Avenue in the City of Bellevue to assess the impact on bicycles and vehicles. Results from the before-and-after study showed that most bicycles used the bike lane and its introduction had little impact on the traffic volume.
- The case study demonstrated the capability of the video analytics platform for both live videos as well as historical videos.
- Microsoft developed a crowd-sourcing platform for labeling the video frames with objects of different categories (vehicles, pedestrians, bicycles, etc.) along with their trajectories. This data is useful for training deep neural networks and computer vision modules.
- Outreach: We built a consortium program on "Video Analytics for Vision Zero" that included major cities in US and Canada.¹ Our work was recognized with two major awards, "Safer Cities, Safer People" US Department of Transportation Award and Institute of Transportation Engineering 2017 Achievements Award.

¹ https://bellevuewa.gov/sites/default/files/media/pdf_document/video-analytics-toward-vision-zero-ITE-Journal-article-March-2017.pdf

Background

In recognition of the opportunities to enhance traffic operations and public safety, Microsoft Corp. and City of Bellevue, entered into a technology development partnership (Figure 1). The collaboration leveraged the city's existing traffic cameras to generate count reports that classify vehicles by turning movement (through, left or right), by direction of approach (northbound, southbound, etc.) and by mode (car, bus, motorcycle, truck, bicycle, pedestrian). From a transportation safety perspective, data on the number of road users differentiated by mode of travel passing through an intersection provides valuable [exposure insights](#) on the relative safety of different locations in a transportation network. In addition to data on the type and motion of road users at intersections, speed and derivatives of speed (e.g., acceleration and sudden motion) can be calculated continuously to better understand steering and braking behaviors. This data has the potential to identify near-collision events, such as when a car abruptly stops or swerves to avoid striking a pedestrian. These close calls are much more frequent and more useful than actual crash reports in detecting systemic safety problems.



Leverage a city's existing traffic camera system to simultaneously:

- **monitor counts and travel speed of all road user groups (vehicle, pedestrian, and bicycle);**
- **document the directional volume of all road user groups as they move through an intersection; and,**
- **assess unsafe "near-miss" trajectories and interactions between all road user groups.**

Figure 1: Video Analytics platform objectives & roadmap.

Data Sources

The City of Bellevue provided Microsoft with access to a sample of live video data. Videos were provided in two forms: pre-recorded videos as well as URLs to live video streams. Pre-recorded videos were uploaded to OneDrive by the City and copied over by Microsoft onto secure servers. For accessing the live video streams, Microsoft utilized a virtual private network (VPN) to log into the City's traffic camera network. For this purpose, Microsoft placed its servers at the "boundary" of its corporate network, and securely accessed the VPN using pre-provided credentials. To ensure reliability, all the cameras were reset every day so as to ensure continuous availability of the live video streams.

Ground Truths

Our case study sought out to obtain the volumes of the multi-modal users – vehicles, bicycles, pedestrians – on the different streets of the City of Bellevue. To compare the accuracy of our techniques, we needed ground truths of the counts. Further, for robustness, we needed these ground truths to cover multiple dimensions that affect video analytics including day and night, different lighting conditions, weather (rainy, sunny days), and different traffic volumes (busy times as well as light traffic times). Microsoft obtained its ground truth from multiple sources. First, ground truths were obtained using extensive labeling by Microsoft’s internal crowd-sourcing platform that allowed for videos to be annotated by users that respect the privacy of content in the videos. Second, Microsoft obtained counts from the City of Bellevue’s inductive in-pavement loops (at 15-minute granularity). It is to be noted that the counts from the loops were available for all the lanes but not necessarily for all the *directions* of movement (e.g., if a lane is meant for both going straight and turning right, these are not differentiated). Finally, Microsoft obtained crowd-sourced data from the public for labeling (Figure 2 shows a screenshot of the public annotation tool). Microsoft compared the accuracy (under-counts, over-counts) of the video analytics system with the above ground truths.

The screenshot displays a web interface for a crowd-sourced video annotation tool. The main heading is "Video Analytics towards Vision Zero". Below this, there are two columns of content. The left column is titled "Worldwide problems demands bold action" and features a video player with a play button. The video player shows a man, identified as Dr. Victor Bahl, Distinguished Scientist, Director, Mobility & Networking, Microsoft Research. Below the video player, there are three bullet points: "Worldwide 1.25 million people are killed annually in traffic accidents", "In 2016, road crashes resulted in 40,000 deaths and 4.6 million injuries in the United States.", and "Crashes are preventable and we need not wait for someone to be killed or injured before we take action". The right column is titled "Make a difference, teach computers to learn" and features a video player with a play button. The video player shows a 3D rendering of a street scene with a car, a cyclist, and a pedestrian. Below the video player, there are three bullet points: "Unique opportunity to help prevent traffic crashes and save lives", "Teach our computers how to recognize vehicles, people walking and bicyclists", and "Cities will be able to rapidly detect road conflicts and traffic engineers can then take preventative action to avoid crashes". At the bottom of the page, there is a blue button with the text "Participate starting June" and a right-pointing arrow.

Figure 2: Screenshot of crowd-sourced video annotation (hosted by ITE).

Rocket Video Analytics System

Microsoft’s [Rocket video analytics system](#) was used to analyze the videos (see Figure 3). Rocket is a highly extensible software stack to empower everyone to build practical real-world live video

analytics applications for object detection and alerting with cutting edge machine learning algorithms. A brief summary of the Rocket platform can be found at <https://aka.ms/Microsoft-Rocket-Video-Analytics-Platform-Rocket-features-and-pipelines.pdf> while the code can be obtained here: <https://github.com/microsoft/Microsoft-Rocket-Video-Analytics-Platform>.

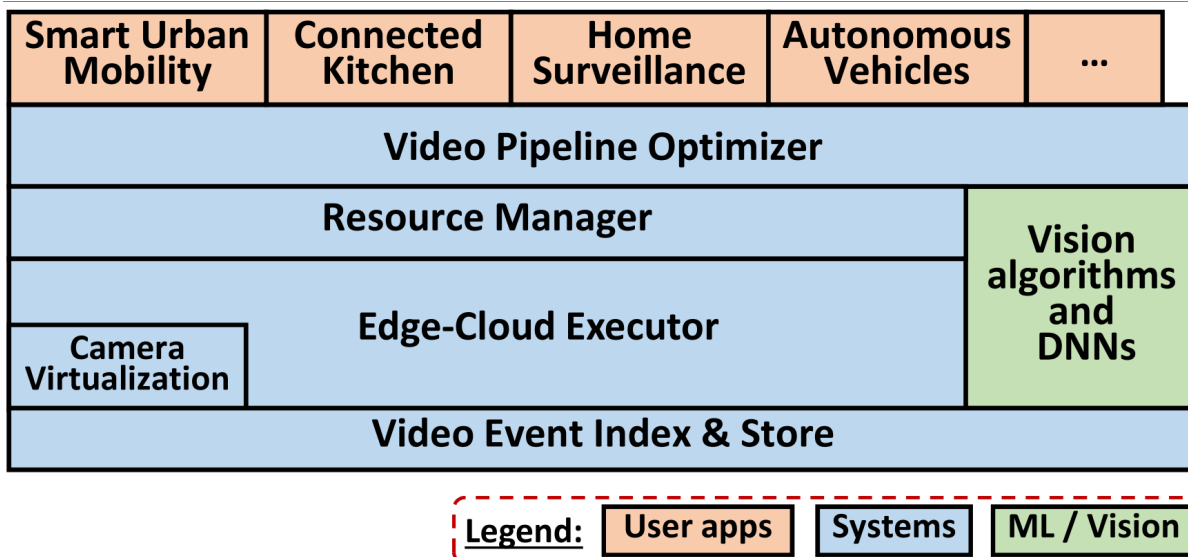


Figure 3: Rocket video analytics software stack. <http://aka.ms/rocket>

The Rocket system (referred to as “system” henceforth) supports multiple applications, with the “queries” of these applications represented as a pipeline of vision modules. Figure 4 shows an example video analytics pipeline that Microsoft used for this study. Intrinsic to the pipeline in Figure 4 is a cascade of operators with increasing cost. The background subtraction module detects changes in each frame and can be run even on CPUs at full frame rate of HD videos. If this module notices a change in the region of interest, it invokes a lightweight DNN model (e.g., tiny Yolo [8]) that checks if there is indeed an object of the queried type (e.g., we may be looking only for cars). Only if the lightweight DNN model does not have enough confidence does Microsoft invoke the heavy DNN model (e.g., full YoloV3 [8]). Such cascading leads to judicious usage of the expensive resources like GPUs.

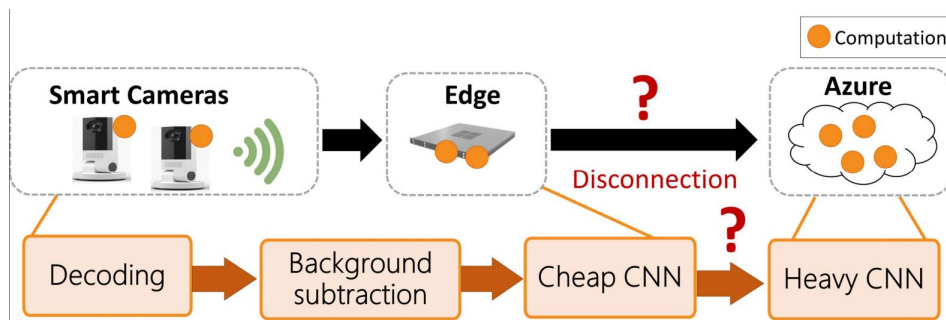



Figure 4: Video analytics pipeline with cascaded operators.



The video pipeline optimizer converts high-level video queries to video-processing pipelines composed of many vision modules; for example, video decoder, followed by object detector and then an object tracker. Each module implements pre-defined interfaces to receive and process events (or data) and send its results downstream. The modules also specify their configuration knobs that Microsoft dynamically modified as part of resource management. The pipeline optimizer also estimates the query's resource-accuracy profile. Specifically, for the different configurations of knobs and implementations of each vision module in the pipeline, it computes the total resource cost and accuracy of the query. To obtain labeled data required to compute accuracy, the optimizer invokes crowdsourcing. The pipeline and its generated profile is then submitted to the resource manager.

Video storage is a key component of the stack and Rocket enables fast and inexpensive retrieval of results from stored videos. Microsoft piggybacks on the live video analytics to use its results as an index for after-the-fact interactive querying on stored videos. Specifically, Microsoft supports asks of the form, find frames with red car in the last week. Microsoft answers such asks without processing a week's volume of videos because the live video analytics allows for the generation of an index of frames in which objects (e.g., red car) occur.

Line-based object counter

Atop the Rocket video analytics system, Microsoft built a line-based technique for uniquely counting objects. The line-based technique is general and allows for obtaining counts of vehicles at the granularity of individual lanes (see Figure 5).

The key aspect of the line-counter is that it captures the "state transition" of the lines, i.e., the state of the line changes from unoccupied to occupied, and then back to unoccupied, before it increases the count for said line. Such a technique is robust to vehicle speeds, vehicle stoppages, as well as addressing vehicles driving close to each other. The lane-based counts are aggregated to obtain directional counts (e.g., northbound turning left), which are provided to the data visualization dashboards.



Figure 5: Line-based object counting.

Data Visualizations

Microsoft created a variety of data visualization dashboards to support the case study objectives. Initially, mock visualizations were created that allowed the City of Bellevue and Microsoft to agree on the objectives of the dashboard. Figure 6 shows how the dashboard will leverage the city’s existing traffic cameras to generate count reports that classify vehicles by turning movement (through, left or right), by direction of approach (northbound, southbound, etc.) and by mode (car, bus, motorcycle, truck, bicycle, pedestrian).

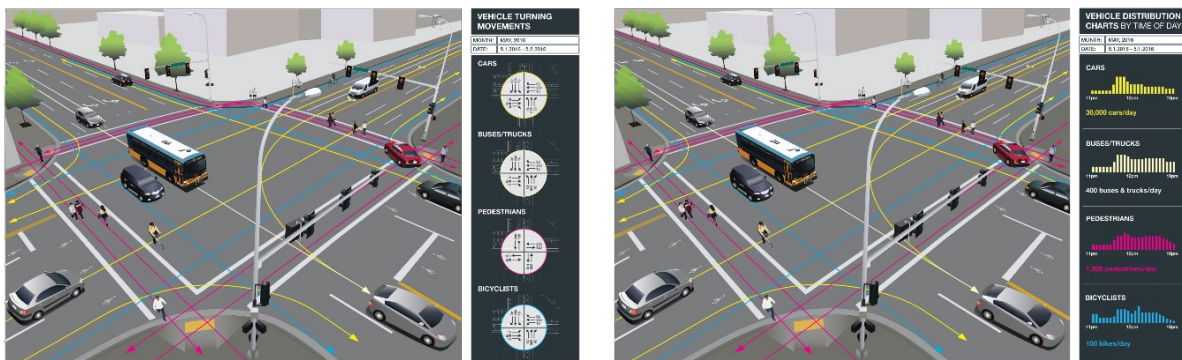


Figure 6: Concept visualization of traffic dashboards

Traffic volume dashboard

Based on the concept visualizations, Microsoft developed a dashboard that ingested the outputs from the video analytics system and populated the dashboard (Figure 7). The dashboard also

provided live alerts when the amount of traffic at an intersection (or in a specific direction) was abnormally high or low; historical traffic volumes were learned to define “normalcy”.

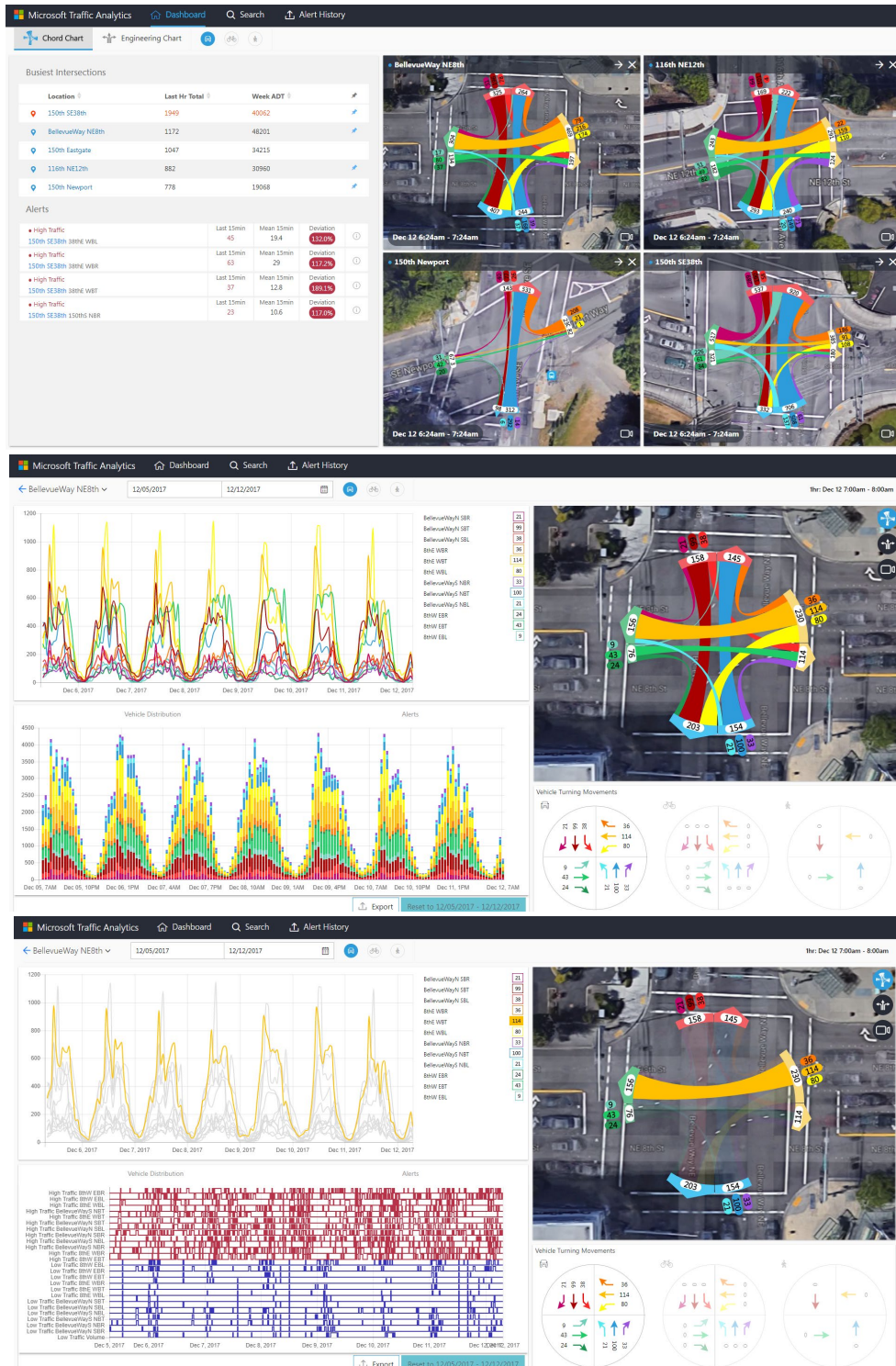


Figure 7: Traffic analytics dashboard deployed at the City of Bellevue’s traffic management center.

The dashboard also provided the ability to analyze traffic volumes over longer periods of multiple months to understand patterns. This help City of Bellevue transportation staff understand the change in traffic volumes, correlate it to documented events (disruptions), and use it for post facto analysis. The user experience, jointly developed and refined with the City of Bellevue, also allowed convenient options to focus on select directions even within an intersection. In addition, staff could also list all known incidents of high as well as low traffic volume events and visualize them in a single screen temporally.

Accuracy of counts: The line-based object detector (described previously) is generic to handle all intersections and directions. Microsoft evaluated its performance across multiple months, to sample and represent different weather conditions (rainy, sunny, cloudy) as well as lighting conditions (day time, evening, nights).

Count accuracies compared to hand-annotated ground truths are positive. The directional counts produced by the video analytics system are 89% to 96% accurate; see Table 1 for accuracies of the five different intersections. Accuracy was measured for individual directions as well as the count for the intersection overall; the latter accuracy was marginally higher.

Camera/Intersection	Accuracy
Bellevue_116th_NE12th	0.9606
Bellevue_150th_Eastgate	0.9464
Bellevue_150th_Newport	0.9426
Bellevue_150th_SE38th	0.9203
Bellevue_BellevueWay	0.8948

Table 1: Accuracy of motor vehicle counts across five intersections in City of Bellevue.

While Microsoft is encouraged by the accuracy of the outputs, it is recognized that further work is required in this space:

- 1) Using a single camera to cover a large intersection invariably results in a (small) fraction of the traffic directions becoming difficult to obtain accurate traffic counts. Causes for this difficulty vary from very few pixels covering this direction or occlusions from other objects. While not insurmountable, these difficulties increase the cost of producing the video analytics output, which in turn negatively affects large-scale adoption. Whenever possible, Microsoft recommends using at least two cameras to cover an intersection.

- 2) Related to the above point is *camera placement*. It is important that cameras are placed in such a way that they cover the approach trajectory of the vehicles as they arrive at the intersection. Doing so provides for inexpensive trajectory-based video analytics techniques.
- 3) Camera resolution and frame rate have a dominant impact on the accuracy of the video analytics. While the video analytics system automatically picks the best choice, it is recommended to invest in cameras with better quality as they lead to accurate outputs.

Bike Lane Planning on 108th Ave

The City of Bellevue built a new bike lane on 108th Avenue to offer the public safe travel options through its Downtown. As part of the planning, the City and Microsoft partnered to assess the interactions of the people riding bicycles and driving cars in the revised lane configuration. See Figure 8 for comparative pictures of the bicycle lane.



Figure 8: Bellevue 108th Avenue, before (left) and after (right) construction of dedicated bike lane.

Since the bicycle lane was constructed by using an existing vehicle travel lane, the objective of the study was two-fold: (a) Are the bike lanes convenient enough for bicycles to use the lane?, and (b) Is the throughput of the vehicular traffic reduced due to one lesser lane?

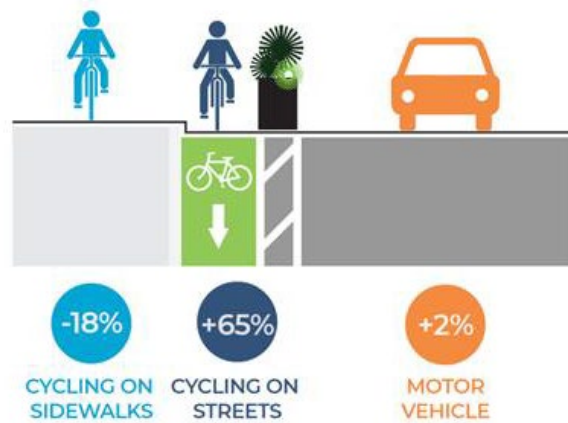


Figure 9: Impact on bicycle and vehicular traffic due to the dedicated bike lane.

The assessment of the results show that the bike lane design met its intended goal. As the graphic (Figure 9) shows, cycling on sidewalks reduced while the use of the bike lane increased. Finally, the throughput of motor vehicles was mostly unchanged.

Precision & Recall of bicycle counts: We also measured the precision and recall of bicycle detections and counts. Precision is defined as the fraction of the bicycles detected by the video analytics system that were in fact bicycles. Recall, on the other hand, is defined as the fraction of bicycles that occurred in the video that the video analytics system detected.

For the three cameras on 108th Avenue intersecting with Main Street, NE 4th Street, and NE 8th Street, we observe recall values of 95% (57 bicycles detected out of the ground-truth of 60), 94% (68 bicycles detected out of the ground-truth of 72), and 91% (60 bicycles detected out of the ground-truth of 66). The precision values, i.e., the number of detected bicycles being bicycles, is over 99%.

Contacts

Microsoft Research: Ganesh Ananthanarayanan (ga@microsoft.com), Victor Bahl, Yuanchao Shu

Microsoft Bing: Peiwei Cao (peiweic@microsoft.com), Fan Xia, Jiangbo Zhang, Ashley Song

City of Bellevue Vision Zero: Franz Loewenherz (FLoewenherz@bellevuewa.gov), Daniel Lai, Darcy Akers

